Canada Border Services Agency   Agence des services frontaliers du Canada

Calibrated (Probabilistic) Confidence Scoring for Biometric Identification

Goal: from scores to probabilities

(0, .5, .5) → (80%, 10%, 10%)

IBPC Conference (NIST), March 2-4, 2010

Dmitry O. Gorodnichy (PhD, CS/Math)
Head of Video Surveillance and Biometrics Section (2008-)

Richard Hoshino (PhD, Math)
Head of Mathematics and Data Exploration Section (2008-)

Science and Engineering Directorate

Canada

CBSA - a prime user of Iris biometrics

Why iris ? – Easily accepted by public, touch-less / non-intrusive

Today: for collaborative user-engaged identification of pre-approved travellers in structured/overt environment (NEXUS)

Tomorrow: for fully-automated stand-off (on-the-fly) identification of Good and Bad people as they cross the border ?(3 persons crossing / sec)

Recent RFI examination (Feb 2009-Aug 2009) exposed the problems even with Today's systems/data

With Tomorrow's stand-off systems, these problems will be even more significant!

Gorodnichy, D. O. *"Evolution and evaluation of biometric systems"* IEEE Symposium: Computational Intelligence for Security and Defence Applications, Ottawa June 2009

Gorodnichy, D. O. *"Multi-order analysis framework for comprehensive biometric performance evaluation"*, SPIE Conf. on Defense, Security, and Sensing. Orlando, April 2010

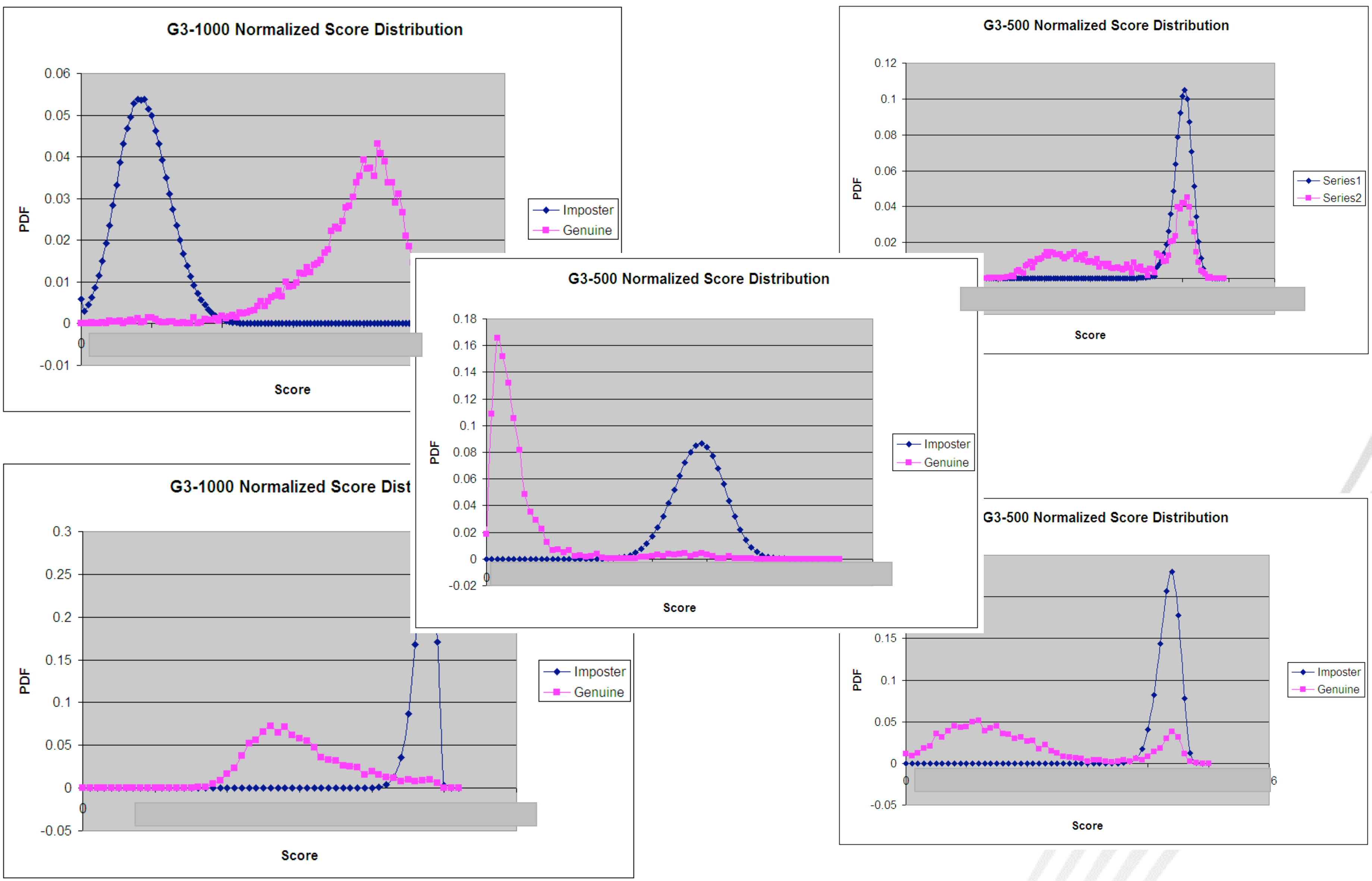
Problems exposed through RFI

(With over 20.000.000 CBSA iris data, several state-of-art products, and over 6 months of coding and collecting/analyzing results)

1. There exist many (>5) matching algorithms now
   - All produce single scores output only (no confidence)!
   - Binomial nature of Imposter distributions
   - Binomial nature of Genuine distribution ? - with no noise
2. High FNMR (False Rejects, False Non-Match Rate)
3. High FTA (Failure To Acquire)
4. Despite many vendor/publications claims, systems <u>often have</u> :
   1) <u>more than one match</u> below the threshold,
   2) two or more <u>close matching scores</u>

<u>There is a need therefore to assign *Confidence value to output*!</u>

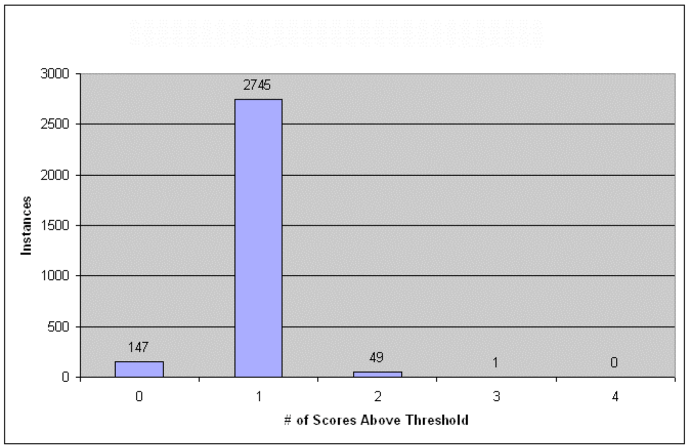
Anonymized score distributions

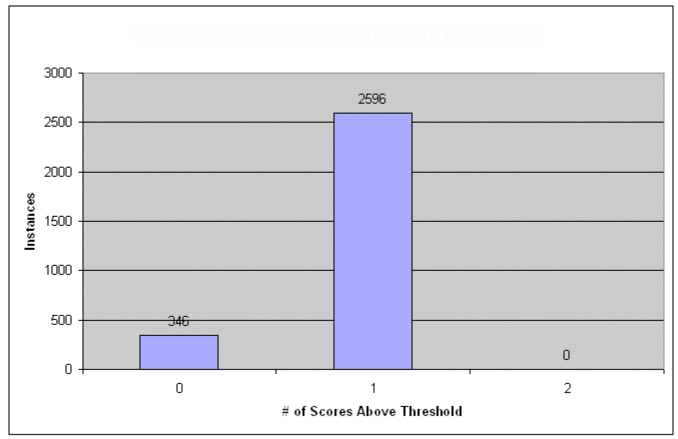


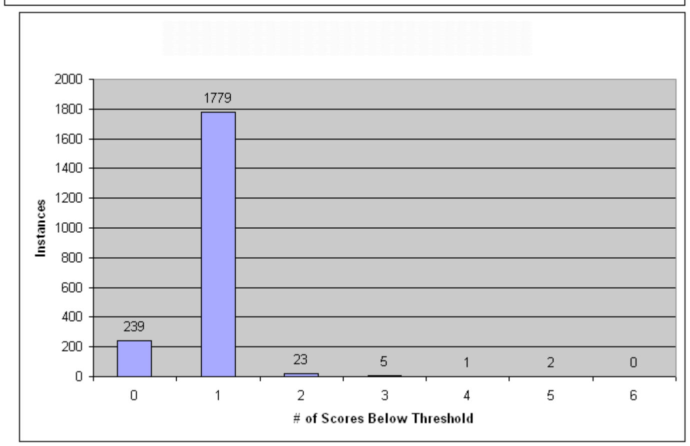
Anonymized stats

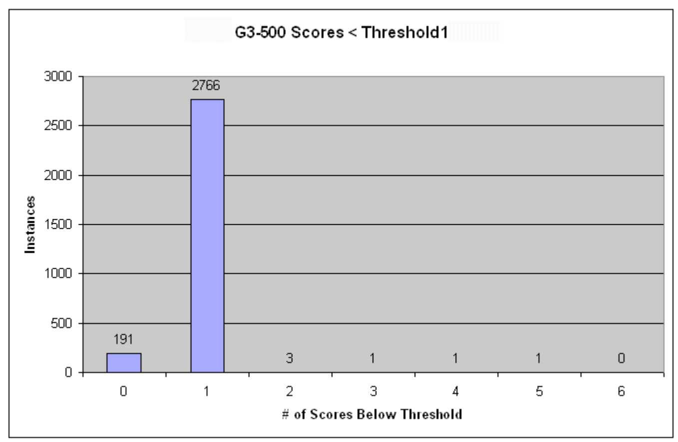
Using Multi-order score analysis [Gor09,10], Order 3 have shown that:

Many systems may improve FTA, FNMR, DET (match/non-match tradeoff) <u>at the cost of allowing more than one score below a threshold</u>





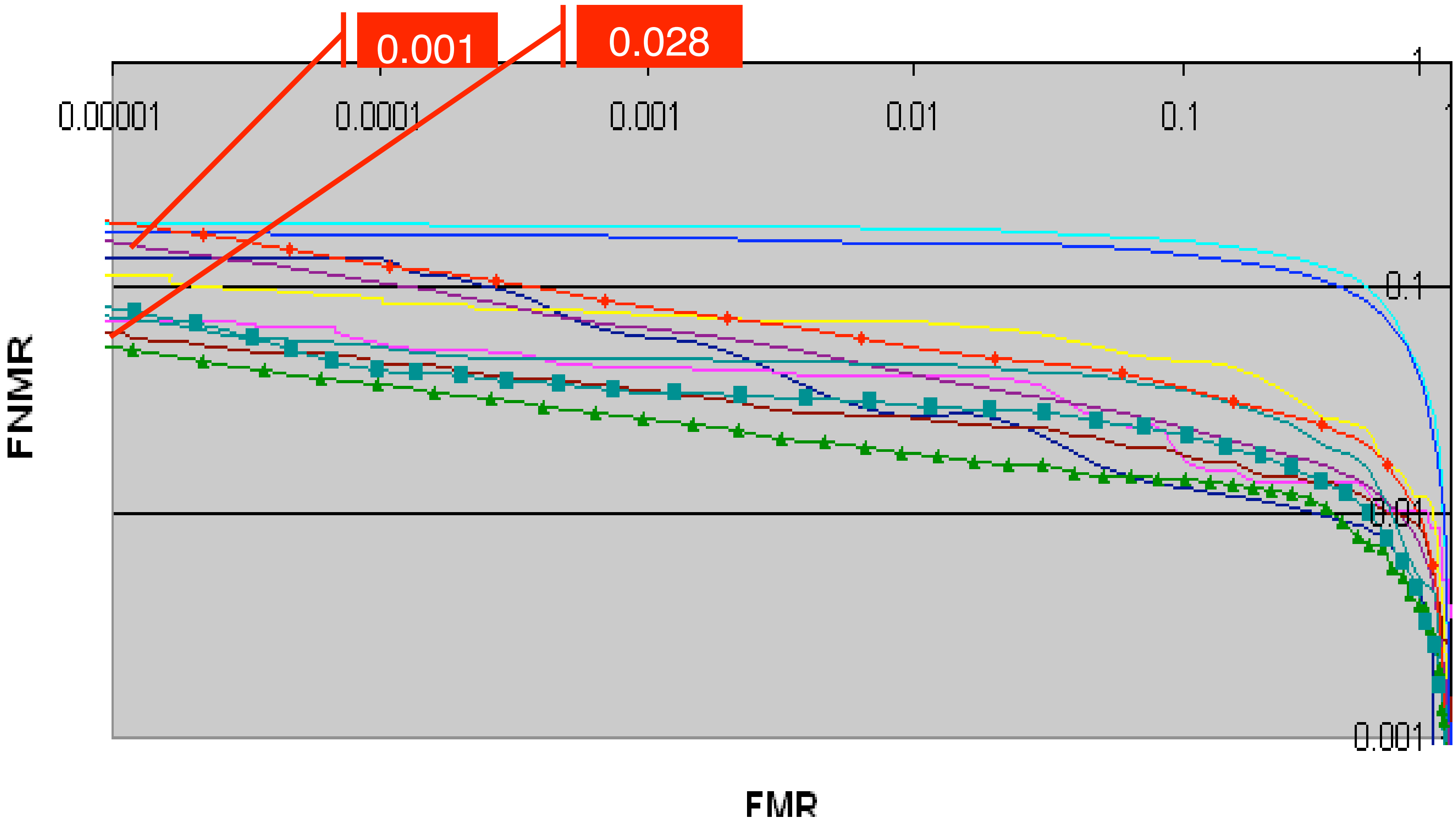




(With 500 enrolled travelers, each having 6 passage images)

Trade-off Curves with FCR

DEFINITION [Gor10]: **Failure of Confidence Rate (FCR)** – the rate of incidences in which there are more than one match below threshold



Goal: assign confidences to decisions

Given: Person X arrives at the kiosk and produces n scores:
n-tuple S = (s1, s2, . . . , sn), si = HD(X, xi)

Find: Sequence of calibrated confidence scores:
the probability vector C = (c1, c2, . . . , cn), ci = P({X = xi} I S)

How: as in probabilistic weather forecasting [DeGroot1983]
1. Make use of (assume) binomial nature of Genuine and Imposter score distributions [Daugman1993,2004]:

- G ~ Binom(m', u'), with u' = 0.11, d' = 0.065 (m'=~115).
- I ~ Binom(m, u), with u = 0.5, m = 249 (d=~0.03)
- P(HD=k/m) = (k,m) u^k (1-u)^(m-k)

2. Bayes's Theorem for ci = P({X = xi} I S) =
= P({X = xi} ΛS) = P({X = xi} Λ S) / P(S) = …

3. P({X = xi} Λ S) = …

Simple example to illustrate

Enrolled: three individuals {x1, x2, x3}, six bits in iris string.

- Thus, n = 3, m = m' = 6.
- G = Binom(m', u'), I = Binom(m, u) with u' = 1/3 and u = 1/2 .
- x1 = [0, 1, 0, 1, 0, 1], x2 = [1, 0, 0, 1, 1, 1], x3 = [1, 0, 1, 1, 0, 1]

New person: X = [0, 1, 0, 1, 0, 1].

- Matching scores S = (0, 0.5, 0.5). Decision scores: (1, 0, 0).

Using the theorem (for q=0 and P1=P2=P3), we obtain:

- confidence scores C = (0.8, 0.1, 0.1).

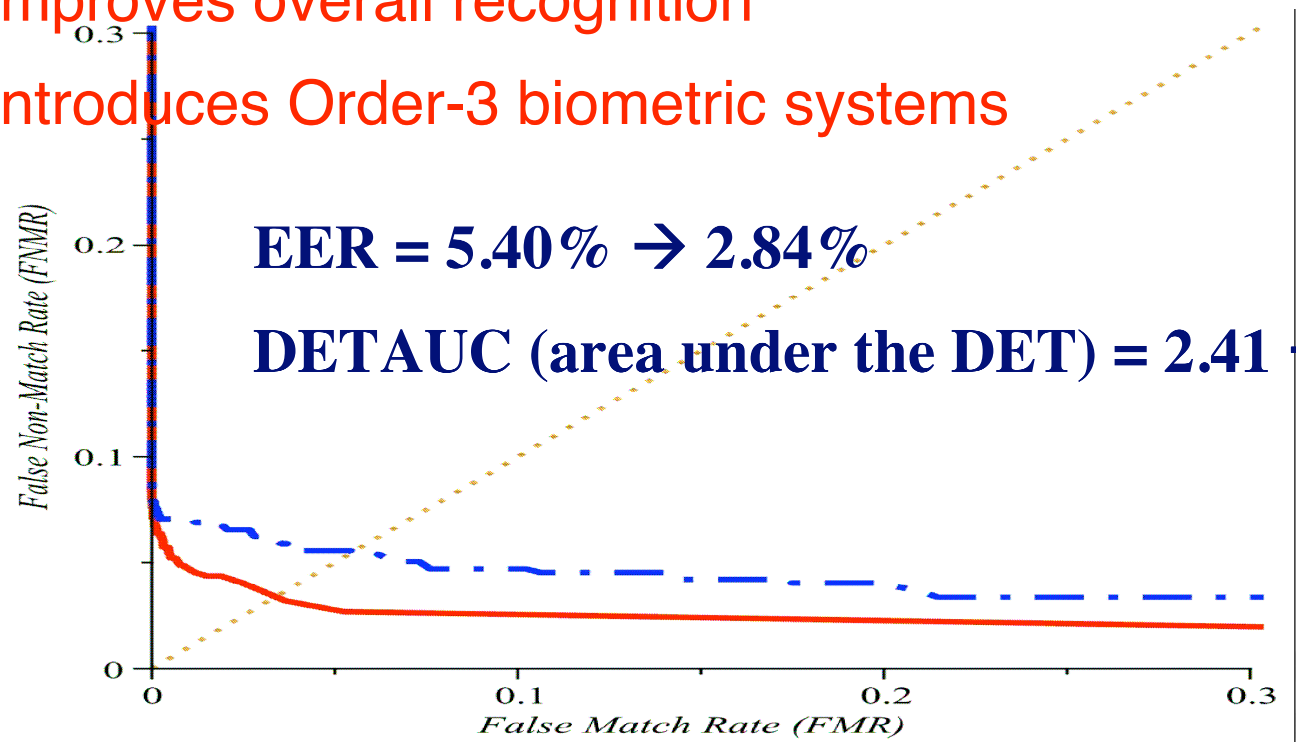
How to apply to real system?

- Vendor should provide: m', u' m, u
- User knows: Pi, q (a-priory probabilities of each person / imposter)

Applied to real system

Proposed probabilistic score calibration can be added to any system at little computation cost as post-processing filter:

- Provides more meaningful output - for risk mitigating procedures
- Improves overall recognition
- Introduces Order-3 biometric systems



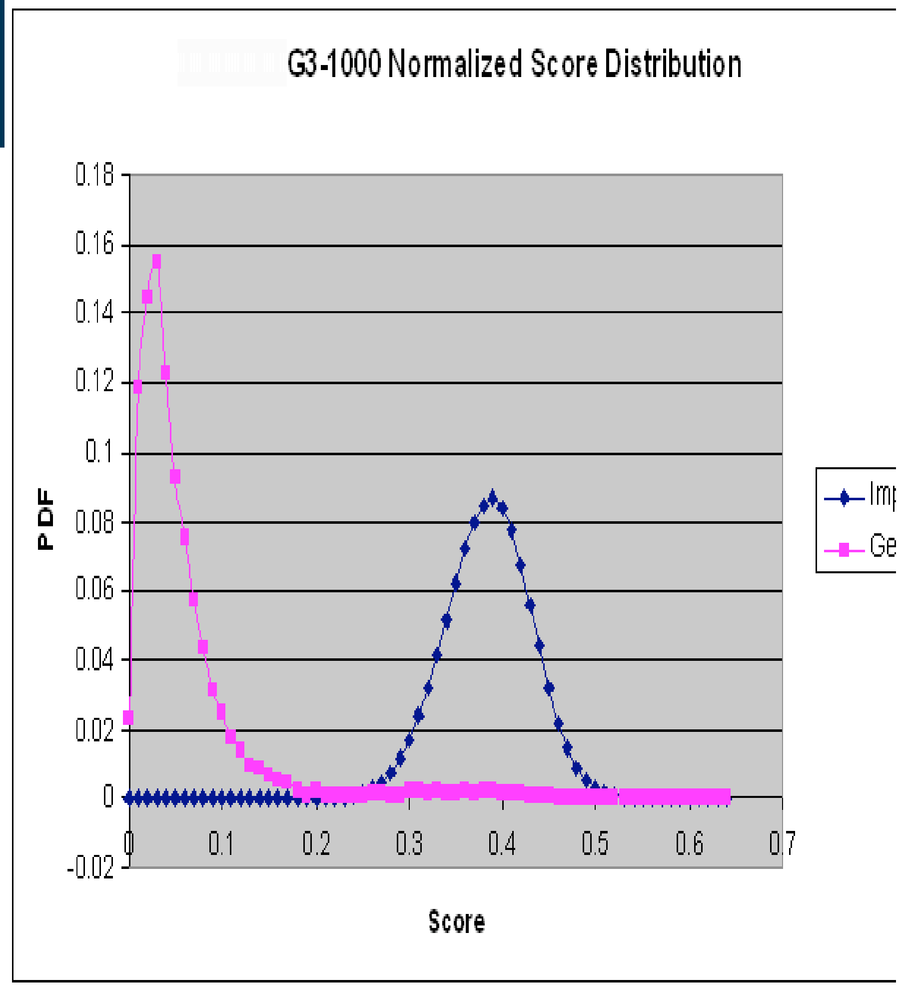
EER = 5.40% → 2.84%

DETAUC (area under the DET) = 2.41 → 0.17

Appendices

Iris biometrics

- Image converted to 2048 binary digits {0,1}
  - only small subsets of bits are mutually independent [1].
- Impostor HD scores follow binomial distribution: I ~ Binom(m, u), m = 249 and u = 0.5.
- The variable m represents the degrees-of-freedom and is a function of the mean u and the standard deviation d: m = u(1 – u) / d^2



- Genuine HD scores [2]: G ~ Binom(m’, u’) with u’ = 0.11, d’ = 0.065

Main theorem and proof:

Theorem 3.1 *Let G be the set of genuine matching scores, and I be the set of impostor matching scores. Suppose $G \sim Binom(\hat{m}, \hat{u})$ and $I \sim Binom(m, u)$. Let $p_i = P(X = x_i)$ and $q = 1 - \sum_{i=1}^{n} p_i$. Let $S = (s_1, s_2, \ldots, s_n)$ be the n-tuple of matching scores produced by person X. Then for each $1 \leq i \leq n$, we have*

$$c_i = P(X = x_i \mid S) = \frac{p_i z_i}{\displaystyle\sum_{i=1}^{n} p_i z_i + q \cdot \frac{(1-u)^m}{(1-\hat{u})^{\hat{m}}}}, \quad \text{where } z_i = \frac{\binom{\hat{m}}{\hat{m}s_i}}{\binom{m}{ms_i}} \cdot \left(\frac{\hat{u}^{\hat{m}}(1-u)^m}{u^m(1-\hat{u})^{\hat{m}}} \right)^{s_i}.$$

Proof: For each $1 \leq i \leq n$, define $r_i = P(\{X = x_i\} \wedge S)$. Also define $r_{imp} = P(\{X \notin \{x_1, x_2, \ldots, x_n\}\} \wedge S)$. By definition, $r_{imp} = P(S) - \sum_{i=1}^{n} r_i$. By Bayes' Theorem, we have

$$c_i = P(\{X = x_i\} \mid S) = \frac{P(\{X = x_i\} \wedge S)}{P(S)} = \frac{r_i}{r_1 + r_2 + \ldots + r_n + r_{imp}}.$$

To calculate $r_i = P(\{X = x_i\} \wedge S)$, we multiply the probabilities of the following $n + 1$ independent events: it is x_i who comes to the kiosk; the genuine matching score $HD(X, x_i)$ is s_i; and the impostor matching score $HD(X, x_j)$ is s_j for all $1 \leq j \leq n$ with $j \neq i$.

Since $G \sim Binom(\hat{m}, \hat{u})$, there are $\hat{m}$ degrees-of-freedom, and the probability that any of these $\hat{m}$ bits differ is $\hat{u}$. So if $HD(X, x_i) = s_i$, then $\hat{m}s_i$ of the $\hat{m}$ bits differ. We derive the analogous result for the impostor distribution $I \sim Binom(m, u)$, for all $1 \leq j \leq n$ with $j \neq i$. Therefore, we have

$$r_i = p_i \binom{\hat{m}}{\hat{m}s_i} \hat{u}^{\hat{m}s_i} (1-\hat{u})^{\hat{m}-\hat{m}s_i} \cdot \prod_{j=1, j\neq i}^{n} \binom{m}{ms_j} u^{ms_j} (1-u)^{m-ms_j}$$

Details of our simple example

Because m = m' = 6, and u = 1-u=1/2, 2*u'=1-u'=2/3 many things get cancelled out …

Zi (Si) = (6, 6*Si) / (6, 6*Si) * ((1/3 ^ 6 * 1/2 ^6) / (1/2 ^ 6 * 2/3 ^ 6)) ^ Si = (1/2^6)^Si = (1/2)^(6*Si)

For S2 = S3 = 0.5, we have: Z2 = Z3 = (1/2)^3 = 1/8.

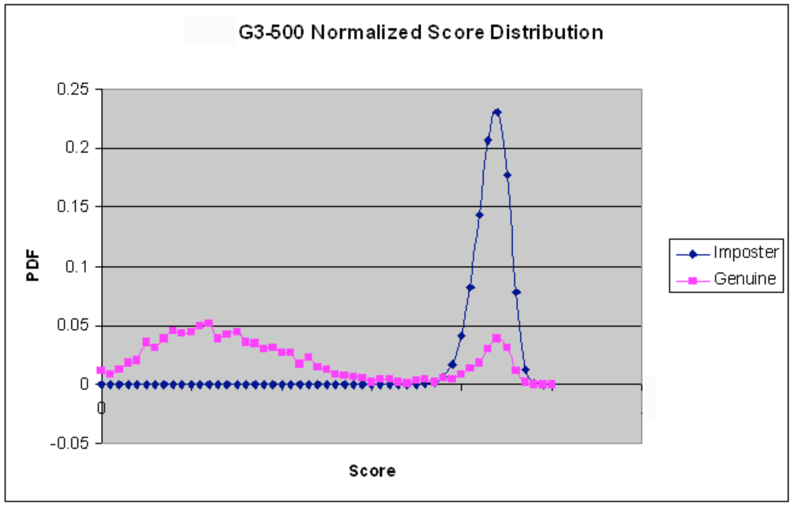
For S1 = 0, Z1 = 1

Then Ci = (Zi) / (SUM Zi) = Zi/ (1/8 + 1/8 + 1) = 4/5* Zi

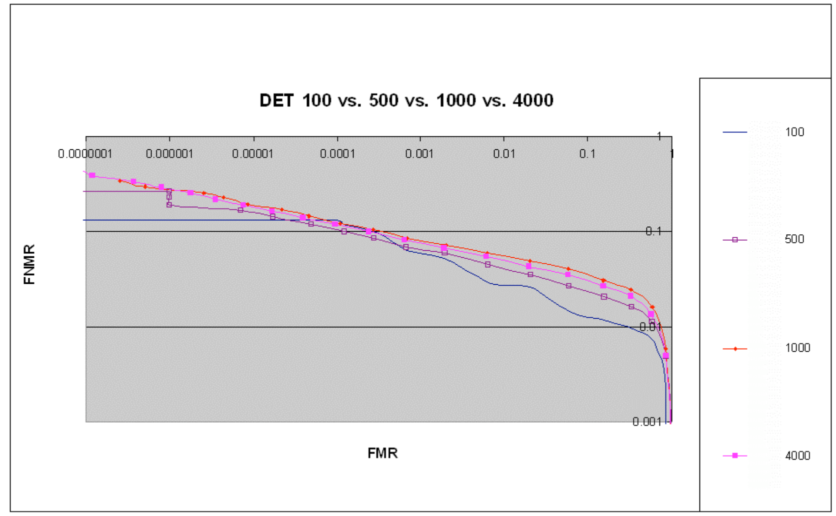
and C2 = 4/5 * (1/8) = 1/10, C1 = 8/10

Multi-order performance evaluation

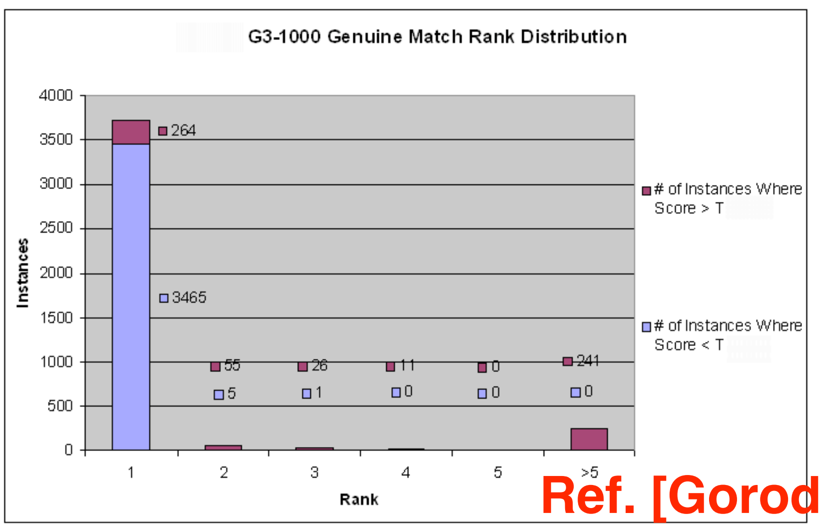
Order 0:



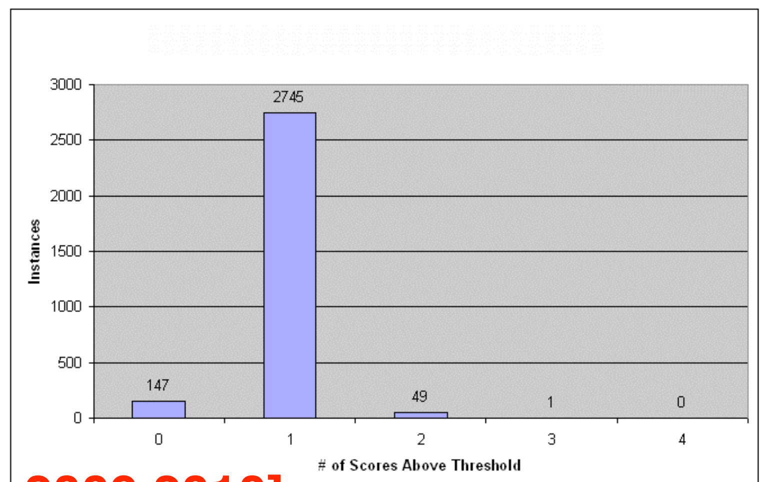
Order 1:



Order 2:



Order 3:



Ref. [Gorodnichy2009,2010]

Multi-order score analysis

Order 1 (Traditional):

- Examine single-scores to report trade-off (FMR/FNMR) curves

Order 2:

- Examine all scores to report the best (smallest) score

Order 3:

- Examine all scores relationship to report Confidences

Five-score example: { 0.51, 0.32, 0.47, 0.34, 0.31 }. T = 0.33

- Order 1 → 0.32
- Order 2 → 0.31
- But in reality it could have been 0.34 ! (if there was noise)

References

Daugman, J. (1993). High confidence visual recognition of persons by a test of statistical independence. IEEE Transactions on Pattern Analysis and Machine Intelligence,

Daugman, J. (2004). How iris recognition works. IEEE Transactions on Circuits and Systems for Video Technology, 14(1) 21-30.

DeGroot, M. H. and Fienberg, S. E. (1983), The comparison and evaluation of probability forecasters, Statistician, 12, 12-22.

Gorodnichy, D. O. (2009). Evolution and evaluation of biometric systems. Proc. of the IEEE Workshop on Applied Computational Intelligence in Biometrics, IEEE Symposium: Computational Intelligence for Security and Defence Applications, Ottawa, 2009.

Gorodnichy, D. O. (2010). Multi-order analysis framework for comprehensive biometric performance evaluation, In Proceedings of SPIE Conference on Defense, Security, and Sensing. DS108: Biometric Technology for Human Identification track. Orlando, 2010